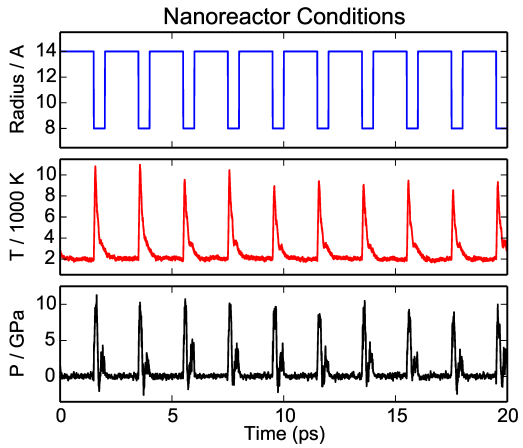
Supplementary Material:
Discovering chemistry with an *ab initio* nanoreactor

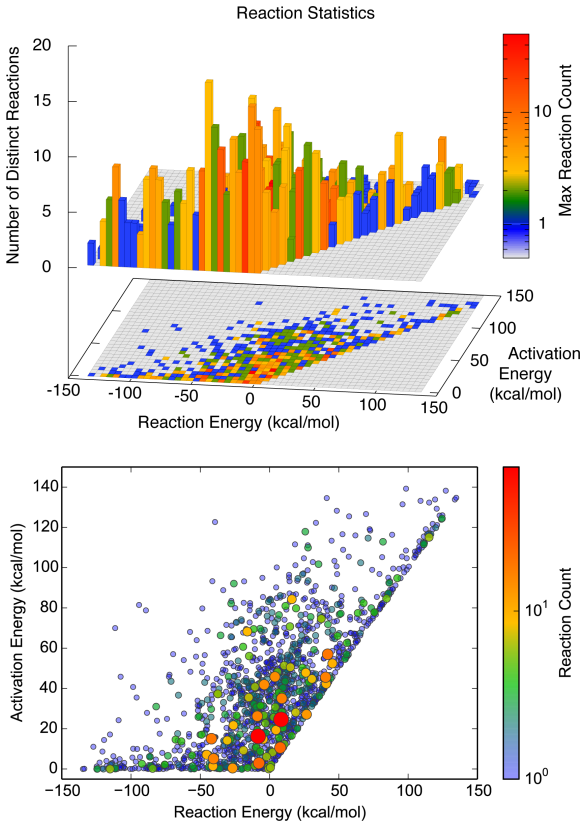
Lee-Ping Wang, Alexey Titov, Robert McGibbon, Fang Liu,
Vijay S. Pande and Todd J. Martínez



Supplementary Figure 1: Time trace of simulation conditions in the first 20 picoseconds of the "Urey-Miller" nanoreactor simulation. *Top*: Radius of the spherical simulation boundary in Angstrom, showing the time-dependent rectangular waveform. *Middle*: Measured simulation temperature calculated from the kinetic energy. The thermostat temperature is 2000 K. *Bottom*: Measured simulation pressure calculated using the virial equation:

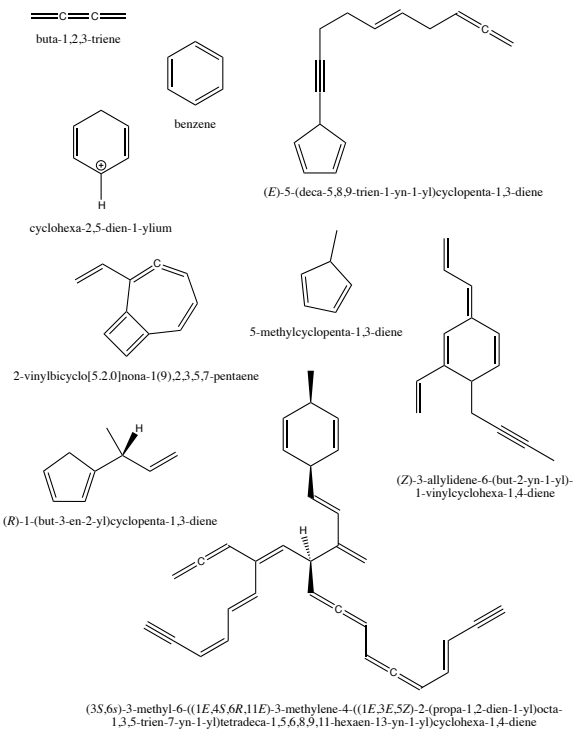
$$PV = Nk_BT + \frac{1}{3}\sum_i \mathbf{r}_i \cdot \mathbf{F}_i \,,$$

where $\mathbf{r}_i$ is the atomic position vector from the sphere center, $\mathbf{F}_i$ is the force component arising from the interatomic interactions, V is the sphere volume, and T is the thermostat temperature. A Hanning window of 51 time steps (25 fs) is used to smooth out high-frequency artifacts in the pressure from bond vibrations. The effect of the sphere radius on the ideal gas part of the pressure is 0.8 GPa, which is small compared to the temporal oscillations.

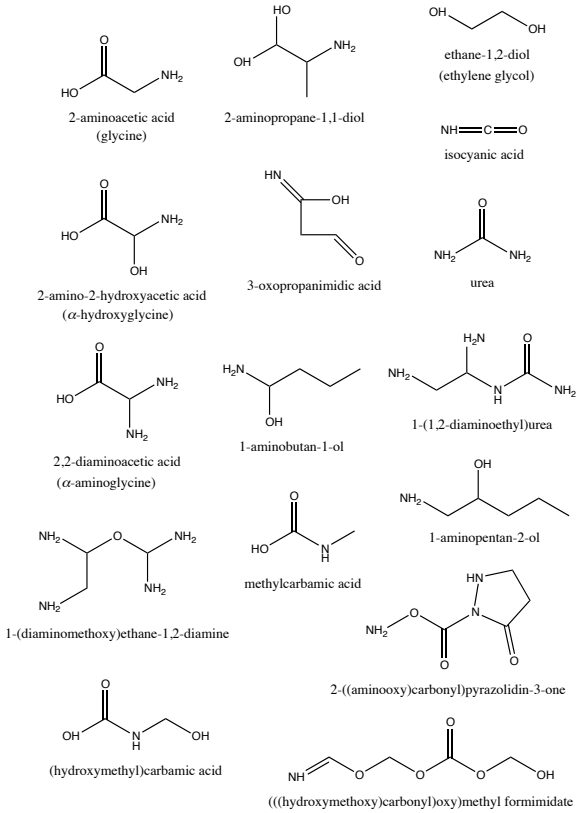


Supplementary Figure 2: Summary of reaction statistics for the "Urey-Miller" nanoreactor simulation.

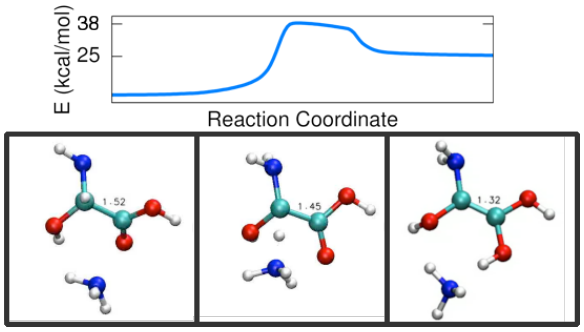
Top: Three-dimensional representation; the bar heights represent the number of distinct reactions, binned by reaction energy and barrier height. The bar colors represent the maximum number of repetitions for any one reaction belonging to that bin. *Bottom*: Two-dimensional representation; each distinct reaction is represented by a circle. The circle size and color corresponds to the number times each distinct reaction was observed.



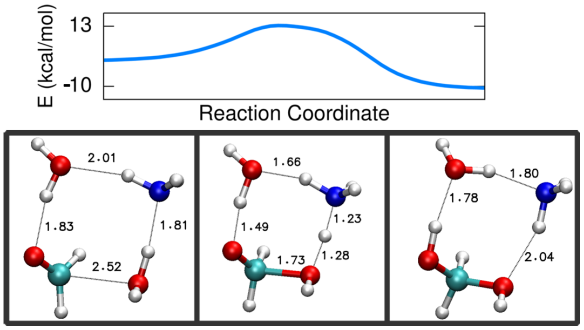
Supplementary Figure 3: A selection of reaction products discovered from the acetylene nanoreactor simulation including linear and branched conjugated chains, allenes, aromatic rings, and a smaller number of highly strained rings.



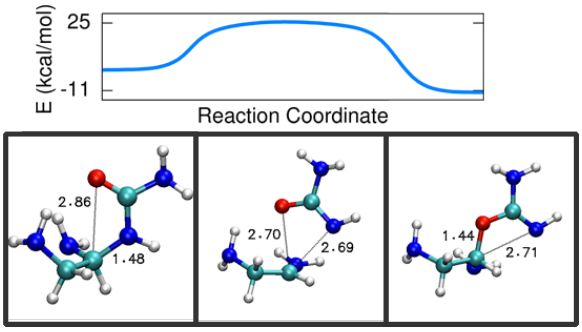
Supplementary Figure 4: A selection of reaction products that were discovered from the "Urey-Miller" nanoreactor simulation including glycine, several unnatural amino acids, and other compounds discussed in the main text. The molecules were chosen to represent the diversity of chemical bonding in the collection of discovered products.



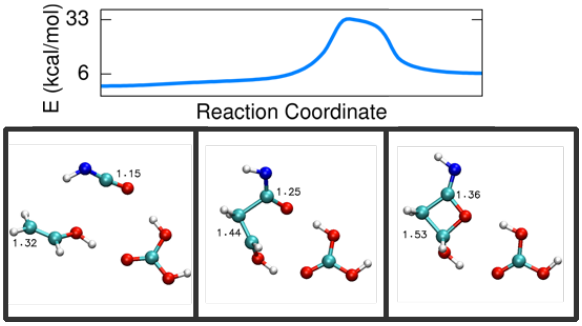
Supplementary Figure 5: Keto-enol isomerization of 2-amino-2-hydroxyacetic acid to 2-aminoethene-1,1,2-triol. This reaction is endothermic by 25 kcal/mol with a 38 kcal/mol barrier. The C-C bond length is marked (Angstroms), showing that the single bond has been transformed to a double bond by the H transfer reaction. In an aqueous environment, one expects that a similar reaction could be catalyzed by a water molecule in place of the ammonia molecule.



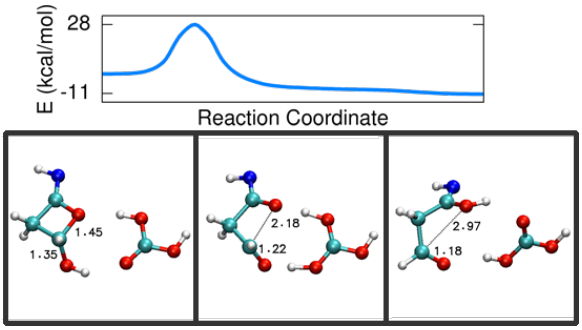
Supplementary Figure 6: Hydration reaction of formaldehyde to methanediol with catalytic water and ammonia molecules. Distances between reacting pairs are labeled (in Angstrom).



Supplementary Figure 7: C-N to C-O bonding rearrangement in 1-(1,2-diaminoethyl) urea.



Supplementary Figure 8: 2+2 addition of isocyanic acid to ethanol, yielding a disubstituted oxetane species.



Supplementary Figure 9: Carbonic acid-catalyzed cleavage of one oxetane C-O bond from the product in the reaction shown in Figure S6 to yield 3-oxopropanimidic acid.

Supplementary Video 1: Animation of nanoreactor simulation on the acetylene system with automatic recognition and highlighting of newly discovered compounds.

Supplementary Video 2: Three-dimensional rotation of Figure 2 showing reaction pyramid in higher detail.